Analog Probabilistic Precipitation Forecasts Using GEFS Reforecasts and Climatology-Calibrated Precipitation Analyses

Thomas M. Hamill,[1] Michael Scheuerer,[2] and Gary T. Bates[2]

[1] *NOAA Earth System Research Lab, Physical Sciences Division, Boulder, Colorado*

[2] *CIRES, University of Colorado, Boulder, Colorado*

Submitted to *Monthly Weather Review*

as an expedited contribution

31 December 2014

Corresponding author:
Dr. Thomas M. Hamill
NOAA Earth System Research Lab
Physical Sciences Division
R/PSD 1, 325 Broadway
Boulder, CO 80305
Tom.Hamill@noaa.gov
Phone: (303) 497-3060
Telefax: (303) 497-6449

ABSTRACT

Analog post-processing methods have previously been applied using precipitation reforecasts and analyses to improve probabilistic forecast skill and reliability. A modification to a previously documented analog procedure is described here that produces highly skillful and statistically reliable precipitation forecast guidance at a somewhat smaller grid spacing. These experimental probabilistic forecast products are available via the web in near real-time.

The main changes to the previously documented analog algorithm were as follows: (a) use of a shorter duration (2002-2013) but smaller grid spacing, higher-quality time series of precipitation analyses for training and forecast verification; (b) increased training sample size using data from 20 locations that were chosen for their similar precipitation analysis climatologies and terrain characteristics; (c) use of point data instead of a set of grid points surrounding a location in determining the analog dates of greatest forecast similarity, and using an analog rather than a rank-analog approach; (d) varying the number of analogs used to estimate probabilities from a smaller number (50) for shorter-lead forecasts to a larger number (200) for longer-lead events; (e) spatial smoothing of the probability fields using a Savitzky-Golay smoother. Special procedures were also applied near coasts and country boundaries to deal with data unavailability outside of the US while smoothing.

The resulting forecasts are much more skillful and reliable than raw ensemble guidance across a range of event thresholds. The forecasts are not nearly as sharp, however. The use of the supplemental locations is shown to especially improve the skill of short-term forecasts during the winter.

1. **Introduction**.

Previous studies have shown that probabilistic forecasts of precipitation can be significantly improved by post-processing with reforecasts (e.g., Hamill et al. 2006, hereafter H06; Hamill et al. 2012, hereafter H12; Hamill and Whitaker 2006, hereafter HW06). The real-time forecast was adjusted using a long time series of past forecasts and associated precipitation analyses. Appealing for its simplicity was the "analog" procedure used therein. For a given location, dates in the past were identified that had reforecasts similar to today's forecast. An ensemble was formed from the observed or analyzed precipitation amounts on the dates of the chosen analogs, and probabilities were estimated from the ensemble relative frequency. Maps of precipitation probabilities were constructed by repeating the procedure across the model grid points.

A challenge with analog procedures used in these previous studies was their inability to find many close-matching forecasts when today's precipitation forecast amount was especially large, even with a long training data set. The method as previously documented used the data surrounding grid point of interest but did not use observation and forecast data centered on other locations. The benefit of this location-specific approach was that if the model's systematic errors varied greatly with location, it corrected for these, as shown in H06. One disadvantage was that if there were not many prior forecasts with similarly extreme precipitation, then the selected analogs were biased toward precipitation forecasts with less extreme forecast values and typically lighter analyzed precipitation. Consequently, the forecast procedure did not often produce high probabilities of extreme events.

Another possible disadvantage of the forecast products demonstrated in these previous studies was that the associated precipitation analyses were in each case from the North American Regional Reanalysis (Mesinger et al. 2006). Several studies have identified deficiencies with this data set (e.g., West et al. 2007, Bukovsky and Karoly 2009). We have also noted a significant dry bias in the NARR over the northern Great Plains during the winter season. There are now alternative data sets covering the contiguous US (CONUS)-based products that utilize both gauge and adjusted radar-reflectivity data. These include the Stage-IV data set (Lin and Mitchell 2005, and http://www.emc.ncep.noaa.gov/mmb/ylin/pcpanl/stage4/) and the climatology calibrated precipitation analysis (CCPA; Hou et al. 2014). Both data sets cover the period of 2002-current. While this time period is shorter than the 1985-current time span of the most recent reforecast (H12), the availability of higher-resolution, more accurate precipitation analysis data has led us to consider whether useful products could be generated with one of these new data sets.

This article briefly describes modifications to previously documented analog forecast procedures. What adjustments will allow it to provide improved probabilistic forecasts while using a shorter time series of analyses? We describe a series of changes to the analog algorithm and show that the resulting analog probabilistic forecasts are skillful and reliable. Since the statistically post-processed guidance provide a significant improvement over probabilities from the raw Global Ensemble Forecast System (GEFS) forecast data, we are also making experimental web-based guidance available in near real time during the next few years; this guidance can be obtained from http://www.esrl.noaa.gov/psd/forecasts/reforecast2/ccpa/index.html.

2. **Methods and data**.

a. *Reforecast data, observational data, and verification methods*.

In this study we will consider 12-hourly accumulated precipitation forecasts during the 2002 to 2013 period for lead times up to +8 days. Precipitation analyses were obtained on a ~1/8-degree grid from the CCPA data set of Hou et al. (2014). Probabilistic forecasts were produced at this ~1/8-degree resolution over the CONUS. All of the forecast data used in this project were obtained from the second-generation GEFS reforecast data set, described in H12. Ensemble-mean precipitation and total-column ensemble-mean precipitable water were used in the analog procedure. GEFS data was extracted (for precipitation) on the GEFS's native Gaussian grid at ~1/2-degree resolution in an area surrounding the CONUS. Precipitable-water forecasts, which were archived on a 1-degree grid, were interpolated to the native Gaussian grid before input to the analog procedure. Forecasts were cross validated; for example, 2002 forecasts were trained using 2003-2013 data.

One of the controls against which the new method was compared were the raw event probabilities generated from the 11-member GEFS reforecast ensemble, bi-linearly interpolated to the 1/8-degree grid.

Verification methods included reliability diagrams and Brier Skill Scores computed in the conventional way (Wilks 2006, eqs. 7.34 and 7.35), with climatology providing the reference probabilistic forecasts. Maps of Brier Skill Scores were also generated for each grid point in the CONUS, accumulating the probabilistic forecasts' and climatological forecasts' average of squared error at that grid point across all years and all months prior to the calculation of skill. Because of the extremely large sample size, confidence intervals for the skill differences (very small; see HW06) were not included on the plots.

b. *Rank analog forecast procedure as a control*.

A "rank analog" approach will serve as another standard for comparison for the newer, somewhat more involved analog methodology described in section 2.c below. For the most part, the rank analog approach is a hybrid of the techniques that have previously been shown to work well, described in sections 3.b.6 and 3.b.8 of HW06. This control rank analog methodology has been further updated in the following respects:

- As with the rank analog algorithm of HW06, the rank of the forecast for a particular date of interest and set of grid points was compared against the ranks of sorted forecasts at the same set of grid points for each date in the training data set. In evaluating which forecasts were closest to today's forecast, the difference between forecasts was calculated as 70% of the absolute difference of the precipitation forecast ranks and 30% of the absolute difference in precipitable water forecast ranks averaged over the set of grid points. Precipitable water was included in the calculation given the slight improvement in warm-season forecasts (HW06) demonstrated from its inclusion.

- The size of the search region for pattern matching of forecasts was allowed to vary with forecast lead time, inspired by the results of testing the method described in 3.b.9 of HW06. Specifically, let t_e denote the end of the forecast precipitation accumulation period in hours, and let δ denote the box width in units of numbers of grid points on the ~ 1/2-degree Gaussian grid. If $t_e \leq 48$, then $\delta=5$; if $48<t_e \leq 96$, then $\delta=7$; if $96<t_e \leq 132$, then $\delta=9$; if $132<t_e$, then $\delta=11$.
- The number of analogs selected was allowed to vary as a function of the forecast lead time and how unusual was the precipitation forecast in question, measured in terms of its percentile relative to the climatological distribution of forecasts (q_f). Let n_a be the number of analogs used. If the end period for the forecast precipitation was > 48 h, then when $q_f<0.75$, $n_a=100$; when $0.75 \leq q_f < 0.9$, $n_a=75$; when $0.9 \leq q_f < 0.95$, $n_a=50$; when $q_f>0.95$, $n_a=25$. If the end period for the forecast ≤ 48 h, then when $q_f<0.75$, $n_a=50$; when $0.75 \leq q_f < 0.9$, $n_a=40$; when $0.9 \leq q_f<0.95$, $n_a=30$; when $q_f>0.95$, $n_a=20$. This dependence of analog size on forecast lead time and unusualness of the forecast with respect to the climatology was inspired by the results of Fig. 7 and associated discussion in H06. This showed that fewer analogs provided the best skill for shorter lead times and for heavy-precipitation events; more analogs were desirable at longer leads and for more common light- or no-precipitation events. The values do not correspond exactly with the optimal values from H06 in part because the length of the training data set is somewhat shorter here.

c. *New analog procedure using data from supplemental locations*.

We now describe an update to the basic analog (hereafter, simply “analog”) procedure described in section 3.a.3 of HW06. This revised procedure will evaluate here and is used in the generation of our real-time web graphics. The following modifications were made:

- Analogs were chosen not by finding a forecast pattern match in an area surrounding the analysis grid point of interest, but rather by using only the forecast data specifically at a grid point. This allowed supplemental data from other grid point locations to be used, uncomplicated by differences of topographic patterns.
- The interpolated forecast for a particular date of interest and analysis grid point (i,j) was compared against interpolated forecasts at (i,j) for each date in the training data set. In evaluating which forecasts were closest to today’s forecast, the difference between forecasts was calculated as 70% of the absolute difference of the precipitation forecasts and 30% of the absolute difference in precipitable water forecasts. Ranks were not compared, as in the prior algorithm, but rather the raw forecasts themselves.
- The interpolated forecast for a particular date of interest and grid point (i,j) was also compared against interpolated forecasts at other supplemental locations (i_s,j_s) on other dates. When a top forecast match was found to occur with data at one of these supplemental locations, then the analysis from this supplemental location on this date was used as an analog member. The first “supplemental” location is merely the original grid point itself. The other 19 supplemental locations were determined for each grid point based upon the similarity of the observed climatology, and the similarity of terrain characteristics.

There were also constraints on a minimum distance between supplemental locations and a penalty for distance between points. The specific methodology of defining supplemental locations is described in the online appendix A. An example of the selected supplemental locations and their dependence on climatology is shown in Fig. 1.

- The number of analogs used in the computation of the probabilities varied with forecast lead time, but not with the unusualness of today's forecast due to the twenty-fold increase in the number of samples. In particular, if the end period t_e for the forecast precipitation was ≤ 24 h, then n_a=50; if 24 < t_e ≤ 48 h, n_a=75; if 48 ≤ t_e < 96 h, n_a=100; if 96 ≤ t_e < 120 h, n_a=150; if t_e ≥ 120 h, n_a=200.

- Once probability forecasts were generated from the ensemble of analyzed states on the dates of the selected forecast analogs, the probability forecasts were smoothed using a 2-D Savitzky-Golay smoother with a window size of 9 grid points and using a third-order polynomial. The details of this smoother are also described in the online appendix A.

3. **Results**.

Figures 2 and 3 show Brier Skill Scores as a function of forecast lead time for the > 1 mm 12 h^{-1} event and the > 95th percentile of climatology event (q95 hereafter), respectively. Skill scores for other event thresholds are presented in online appendix B. While both rank analog and analog forecasts provided a significant improvement with respect to the raw guidance, the skills of the newer analog method for this event were not appreciably different from those of the rank analog method. This was likely because the > 1 mm event was not an especially rare event at most locations, so the increased sample size with the new analog method was not particularly critical. Considering the skill for q95 in Fig. 3, the new analog procedure does provided a skill improvement, especially for shorter-lead forecasts during the cool season. In these circumstances, the day +2 analog forecasts with supplemental locations were comparable in skill to the day +1 rank analog forecasts, and both were dramatically higher in skill than the raw ensemble. Why was there improvement with the new analog procedure in winter? Though not confirmed, we hypothesize that in winter there was higher intrinsic skill of the forecasts than in summer, due to the different phenomena driving precipitation with their different space and time scales: synoptic-scale ascent in mid-latitude winter cyclones, thunderstorms during the summer season. Further, in wintertime, there were larger fluctuations of the probabilities about their long-term climatological mean with meaningful signal. Thus the additional samples helped refine the estimates of O|F, the conditional distribution of observations given the forecast (HW06, eq. 3), thereby improving the probabilistic forecast.

Figure 4 shows maps of Brier skill scores for the > 1 mm event at the 60-72-h lead time. There was little difference between the two analog forecasts, consistent with Fig. 2. Both were more skillful than the raw ensemble, which has BSS < 0 over a significant percentage of the country, in part due to sampling error (Richardson 2001) but mostly due to systematic errors and sub-optimal treatment of model uncertainty in the GEFS. Skill was largest along the US West Coast, with the predictable phenomena of the flow from mid-latitude cyclones impinging upon the stationary topography. Figure 5 shows maps of skill for the > q95 event at the 60-72-h lead time. There were greater differences between the analog with supplemental locations and the rank analog without; there appeared to be a general improvement in skill across the country for the analog with supplemental locations, perhaps enhanced more than average in the rainy areas along the US West Coast. Again, raw ensembles were notably unskillful across drier regions of the US. Maps for other forecast lead times and thresholds are provided in online Appendix B.

The resulting post-processed forecast guidance was consistently reliable, too. Figure 6 provides reliability diagrams for the three methods for > q95 and 60-72 h forecast leads; again, see appendix B for more diagrams at other leads and event thresholds. Both analog methods were quite reliable, though the analog with supplemental locations had somewhat more forecasts issuing high-probabilities. Both analog methods were much less sharp than the raw forecast guidance but more reliable.

4. **Discussion and conclusions**

This article has demonstrated an improved method for post-processing that provides dramatically improved guidance of probabilistic precipitation when paired with a reforecast data set of sufficient length and precipitation analyses of sufficient quality. This article provides additional evidence to support the assertion that the regular production of weather reforecasts will help with the objective definition of high-impact event probabilities.

This method may provide a useful benchmark for comparison of other methods. Whereas the analog method here has been shown to work well with larger reforecast data sets, these are not always available. We anticipate subsequent studies will compare the efficacy of analog methods with respect to other (e.g., parametric) post-processing methods when using much smaller training sample sizes. In this way we hope to understand whether the choice of post-processing algorithm is robust across sample sizes.

Acknowledgments:

This research was supported by a NOAA US Weather Program grant as well as funding from the National Weather Service Sandy Supplemental project. The reforecast data set was computed at the US Department of Energy's (DOE) National Energy Research Computing Center, a DOE Office of Science user facility.

References

Bukovsky, M. S., and D. J. Karoly, 2009: A brief evaluation of precipitation from the North American Regional Reanalysis. *Journal Hydrometeor.,* **8**, 837-846.

Hamill, T. M., J. S. Whitaker, and S. L. Mullen, 2006: Reforecasts, an important dataset for improving weather predictions. *Bull. Amer. Meteor. Soc.*, **87**,33-46.

Hamill, T. M., and J. S. Whitaker, 2006: Probabilistic quantitative precipitation forecasts based on reforecast analogs: theory and application *Mon. Wea. Rev.*, **134**, 3209-3229.

Hamill, T. M., G. T. Bates, J. S. Whitaker, D. R. Murray, M. Fiorino, T. J. Galarneau, Jr., Y. Zhu, and W. Lapenta, 2012: NOAA's second-generation global medium-range ensemble reforecast data set. *Bull Amer. Meteor. Soc.*, **94**, 1553-1565.

Hou, D., M. Charles, Y. Luo, Z. Toth, Y. Zhu, R. Krzysztofowicz, Y. Lin, P. Xie, D.-J. Seo, M. Pena, and B. Cui, 2014: Climatology-calibrated precipitation analysis at fine scales: statistical adjustment of Stage IV toward CPC gauge-based analysis. *J. Hydrometeor*, **15**, 2542–2557. doi: http://dx.doi.org/10.1175/JHM-D-11-0140.1

Lin, Y., and K. E. Mitchell, 2005: The NCEP Stage II/IV hourly precipitation analyses: Development and applications. *19th Conf. on Hydrology*, San Diego, CA, Amer. Meteor. Soc., 1.2. Available online at https://ams.confex.com/ams/Annual2005/techprogram/paper_83847.htm .

Mesinger, F., and others, 2006: North American regional reanalysis. *Bull. Amer. Meteor. Soc.*, **87**, 343-360.

Richardson, D. L., 2001: Measures of skill and value of ensemble prediction systems, their interrelationship and the effect of ensemble size. *Quart. J. Royal Meteor. Soc.*, **127**, 2473-2489.

West, G. L., W. J. Steenburgh, and W. Y. Y. Chen, 2007: Spurious grid-scale precipitation in the North American Regional Reanalysis. *Mon. Wea. Rev.*, **153**, 2168-2184.

Figure captions

Figure 1. Illustration of the location of supplemental locations and their dependence on the analyzed precipitation climatology. Climatology is shown for the 95th percentile of the analysis distribution for the month of January, based on 2002-2013 CCPA data. Supplemental data locations are also shown. The larger symbols indicate sample locations where supplemental data is sought, and the smaller symbols indicate the chosen supplemental locations.

Figure 2: Brier skill scores for the > 1 mm event over a range of lead times as a function of the month of the year. (a) Skills of forecasts from the new analog method; (b) skills of forecasts from the older rank-analog method for comparison; (c) skills of forecasts from the 11-member raw ensemble guidance.

Figure 3: As in Fig. 2, but for the event of greater than the 95th percentile of the climatological analyzed distribution. The climatology is computed separately for each month and each ~1/8-degree grid point location.

Figure 4: Maps of yearly 60-72 h forecast Brier Skill Scores, for probabilistic forecasts of the > 1 mm 12 h^{-1} event, generated from (a) analog forecasts with 20 supplemental locations, (b) rank analog forecast with no supplemental locations, and (c) 11-member raw ensemble.

Figure 5: As in Fig. 4, but for > q95 event.

Figure 6: Reliability diagrams for the > q95 event for 60- to 72-h forecasts. (a) analog forecasts with 20 supplemental locations, (b) rank analog forecast with no supplemental locations, and (c) 11-member raw ensemble.

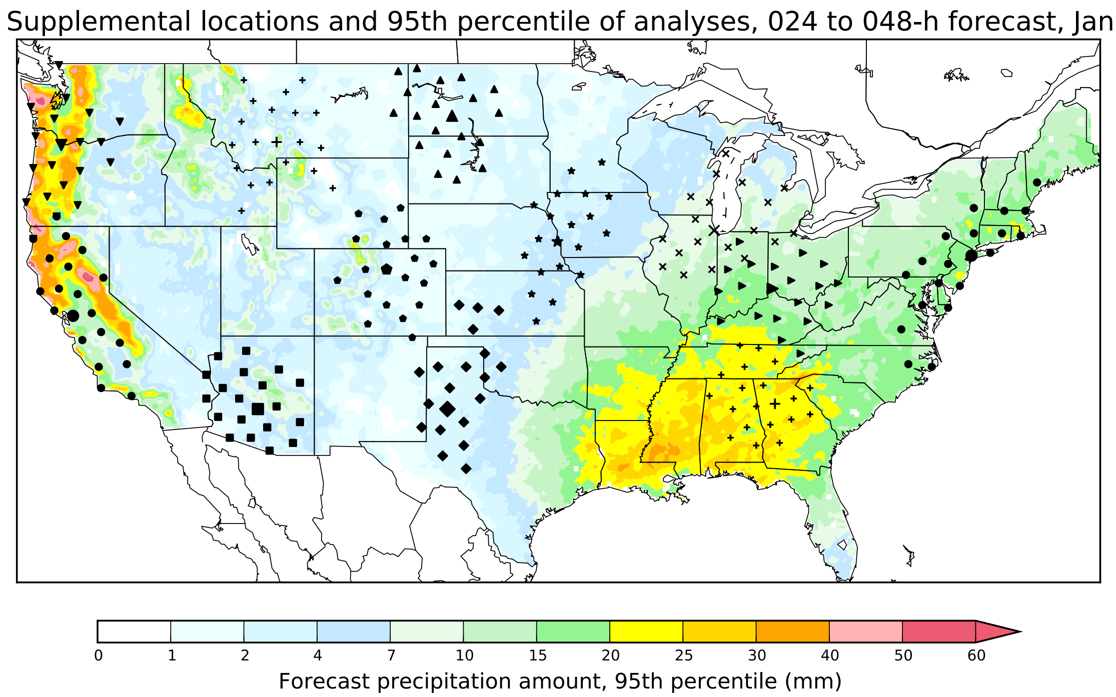

Figure 1. Illustration of the location of supplemental locations and their dependence on the analyzed precipitation climatology. Climatology is shown for the 95th percentile of the analysis distribution for the month of January, based on 2002-2013 CCPA data. Supplemental data locations are also shown. The larger symbols indicate sample locations where supplemental data is sought, and the smaller symbols indicate the chosen supplemental locations.

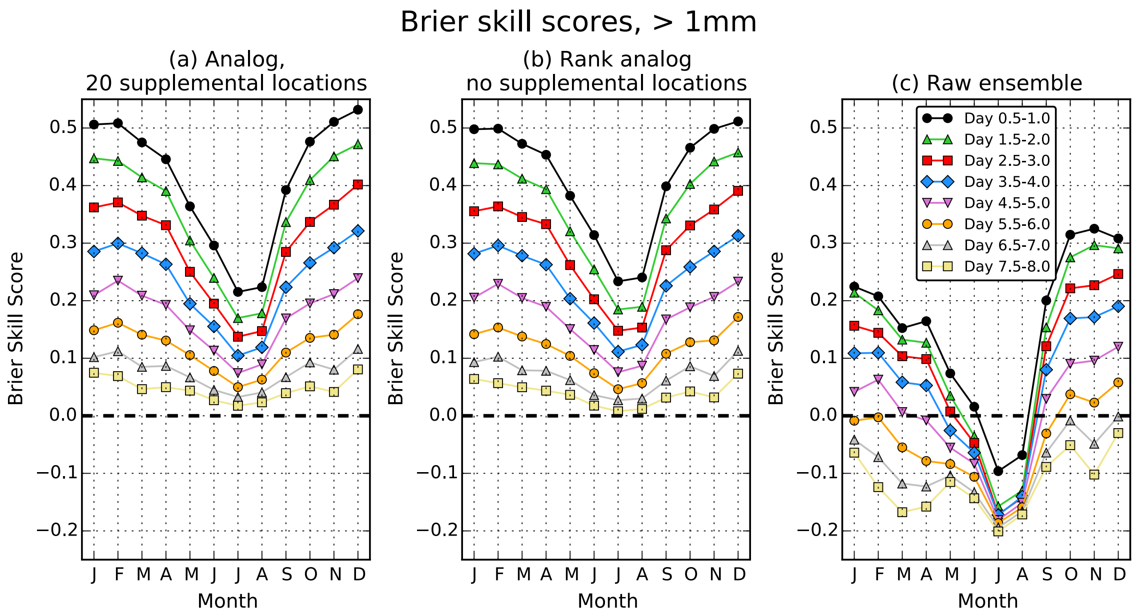

Figure 2: Brier skill scores for the > 1 mm event over a range of lead times as a function of the month of the year. (a) Skills of forecasts from the new analog method; (b) skills of forecasts from the older rank-analog method for comparison; (c) skills of forecasts from the 11-member raw ensemble guidance.

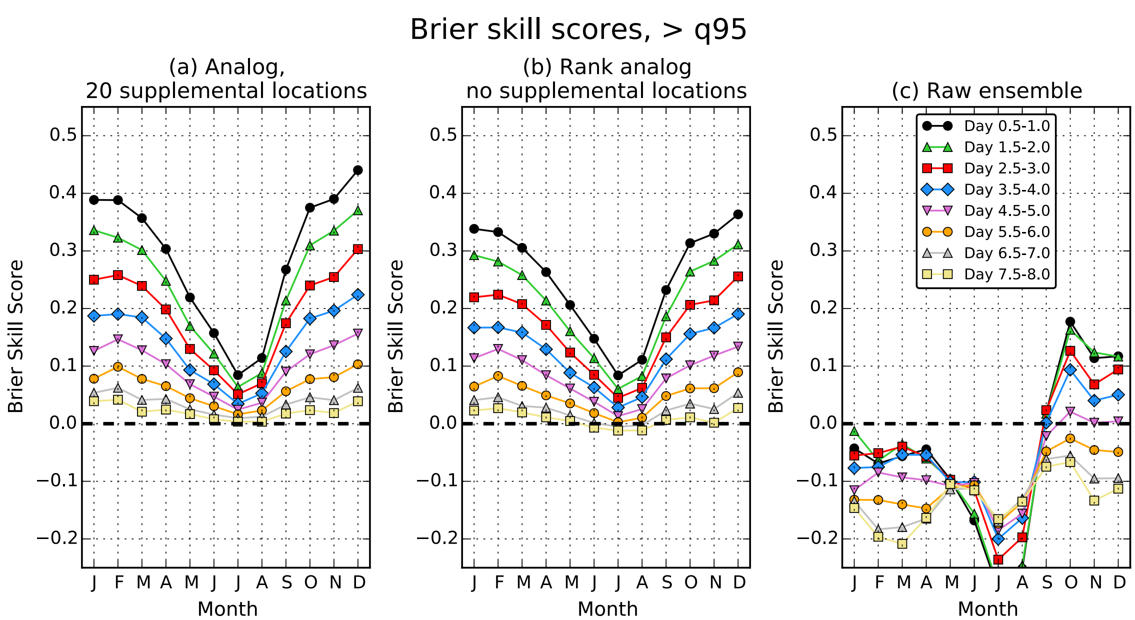

Figure 3: As in Fig. 2, but for the event of greater than the 95th percentile of the climatological analyzed distribution. The climatology is computed separately for each month and each ~1/8-degree grid point location.

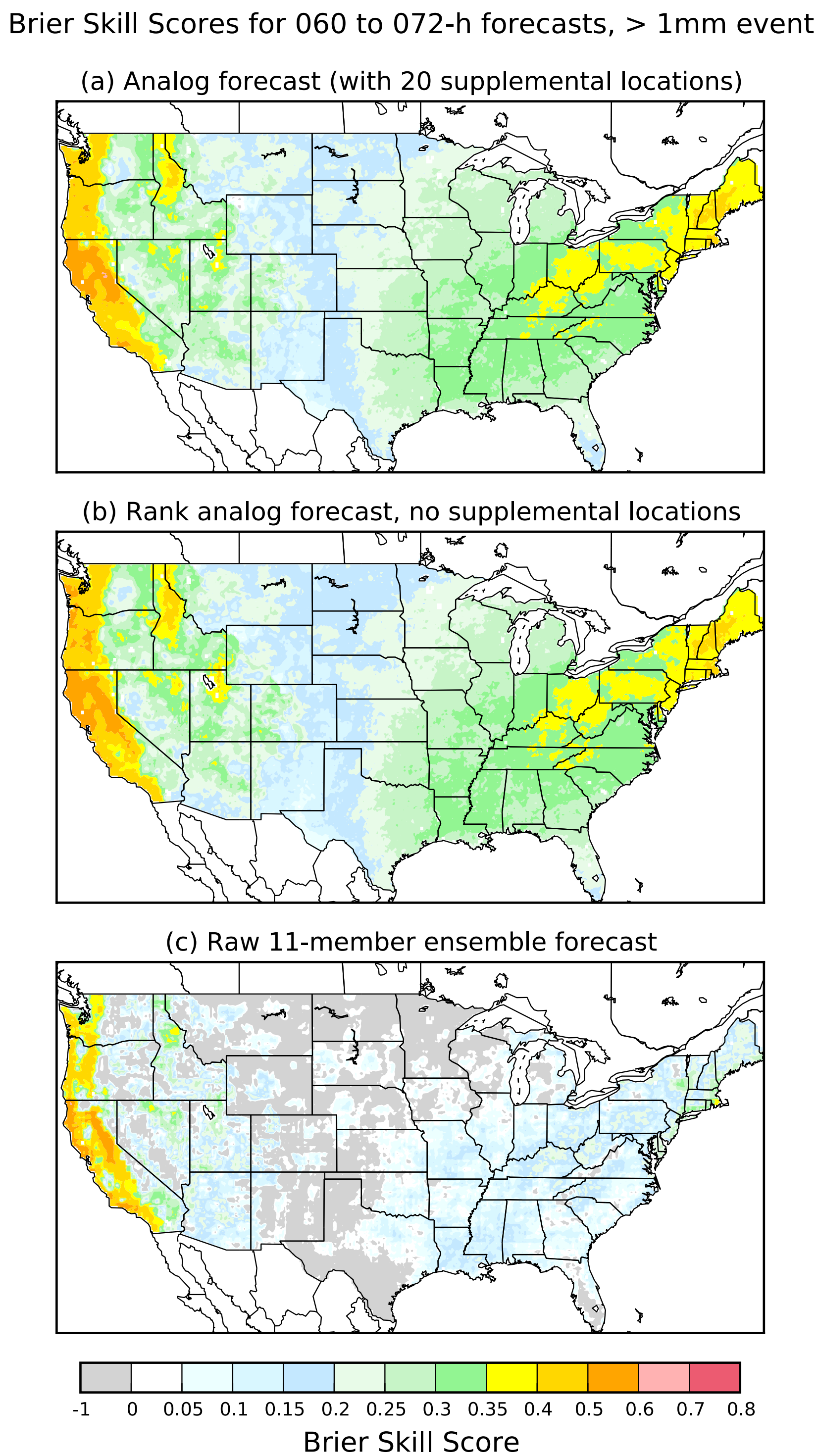

Figure 4: Maps of yearly 60-72 h forecast Brier Skill Scores, for probabilistic forecasts of the > 1 mm 12 h^{-1} event, generated from (a) analog forecasts with 20 supplemental locations, (b) rank analog forecast with no supplemental locations, and (c) 11-member raw ensemble.

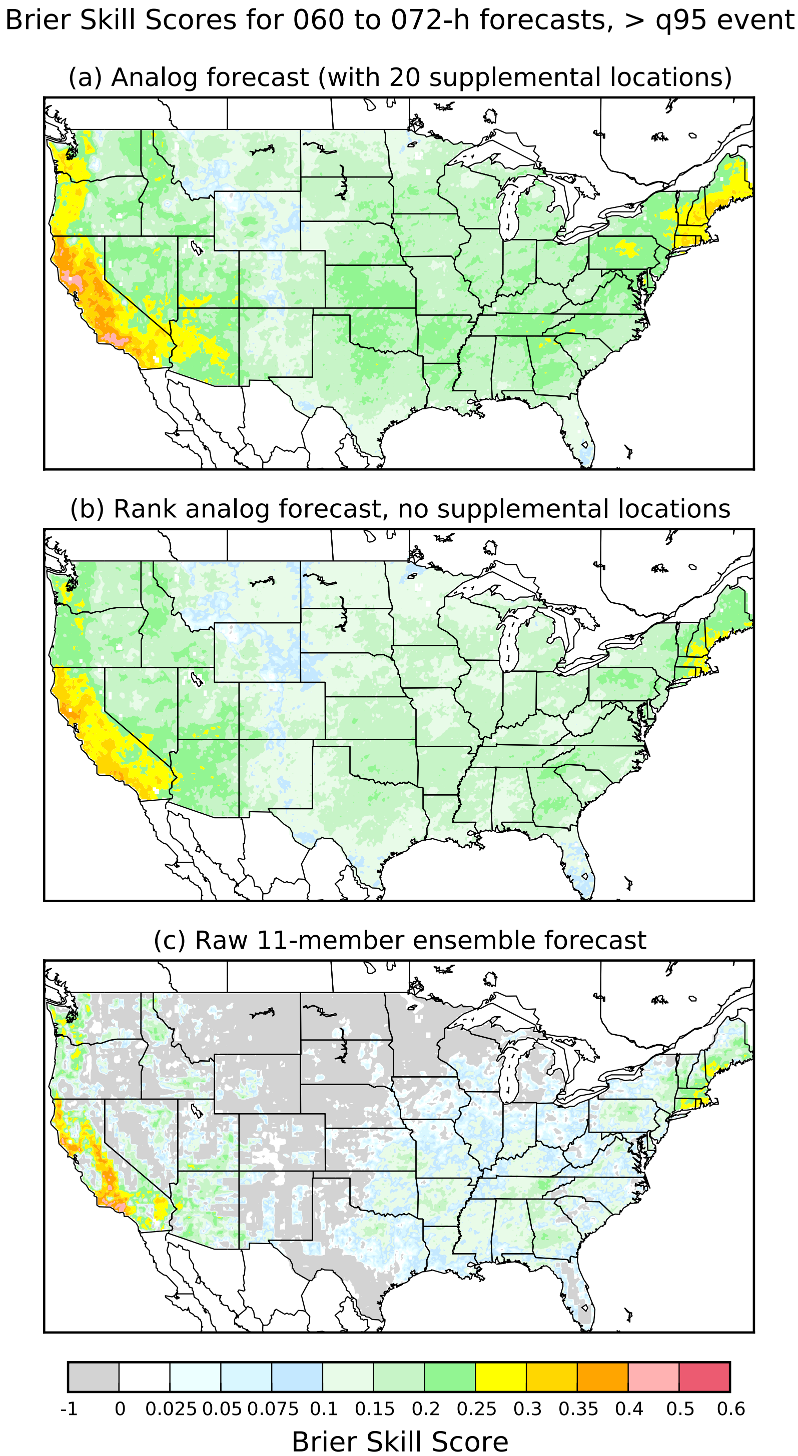

Figure 5: As in Fig. 4, but for > q95 event.

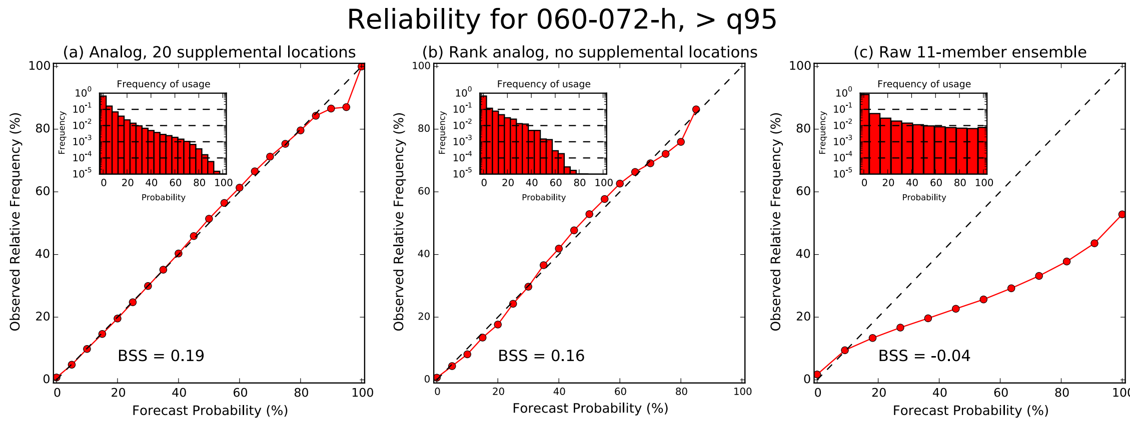

Figure 6: Reliability diagrams for the > q95 event for 60- to 72-h forecasts. (a) analog forecasts with 20 supplemental locations, (b) rank analog forecast with no supplemental locations, and (c) 11-member raw ensemble.